Toward Less Misleading Comparisons of Uncertain Risks: The Example of Aflatoxin and Alar

Adam M. Finkel

Cecil and Ida Green Center for the Study of Sciences and Society, University of Texas, Richardson, TX 75080 USA

Critics of comparative risk assessment (CRA), the increasingly common practice of juxtaposing disparate risks for the purpose of declaring which one is the "larger" or the "more important," have long focused their concern on the difficulties in accommodating the qualitative differences among risks. To be sure, people may disagree vehemently about whether "larger" necessarily implies "more serious," but the attention to this aspect of CRA presupposes that science can in fact discern which of two risks has the larger statistical magnitude. This assumption, encouraged by the indiscriminate calculation of risk ratios using arbitrary point estimates, is often incorrect: the fact that environmental and health risks differ in unknown quantitative respects is at least as important a caution to CRA as the fact that risks differ in known qualitative ways. To show how misleading CRA can be when uncertainty is ignored, this article revisits the claim that aflatoxin contamination of peanut butter was "18 times worse" than Alar contamination of apple juice. Using Monte Carlo simulation, the number 18 is shown to lie within a distribution of plausible risk ratios that ranges from nearly 400:1 in favor of aflatoxin to nearly 40:1 in the opposite direction. The analysis also shows that the "best estimates" of the relative risk of aflatoxin to Alar are much closer to 1:1 than to 18:1. The implications of these findings for risk communication and individual and societal decision-making are discussed, with an eye toward improving the general practice of CRA while acknowledging that its outputs are uncertain, rather than abandoning it for the wrong reasons. *Key words*: carcinogenic risk, HERP index, natural and synthetic pesticides, risk comparison, uncertainty. *Environ Health Perspect* 103: 376–385 (1995)

To assess risk is to compare risks. Comparisons are hidden or overt virtually any time data and models are used to quantify some environmental or health hazard. This holds true whether the social purpose involves setting a standard (which entails comparing the risk without any intervention to the magnitude, uncertainty, and distribution of risks after intervening), communicating the findings of science (disembodied risk estimates are meaningless to most people without reference to background rates or other numerical indices), or setting priorities (without comparisons, either nothing would be a priority or, equivalently, everything would). And yet, against the countless person-years of effort that have gone into refining and codifying the methodology for quantifying one risk at a time, there has been virtually no progress in developing principles and methods for quantifying risk comparisons.

Comparative risk assessment (CRA) is too important to do poorly. Not only do government agencies use CRA to influence the way people think about different risks, but they are increasingly using it to make irrevocable choices about which risks to control and which to accept. Government must decide, for example, whether to promote, mandate, or restrict alternative fuels such as methyl tert-butyl ether (MTBE) for automobiles; its only choice is whether to use CRA to compare gasoline and MTBE or instead to make the decision on intuitive, political, or other grounds. Either way, choices such as these will be made, but reliance on a misleading analytic tool might be worse than undertaking no analysis at all.

At its current state of development, however, CRA may be sufficiently flawed that on balance it causes more harm than good. Decision-makers cannot use CRA without asking whether merely knowing which of two risks is statistically larger is sufficient to guide regulatory policy or individual choice. Even putting this aside, however, there remains a purely scientific question: With current methods of CRA, would we know a "larger" risk when we saw it?

This article explores a largely unrecognized but fundamental flaw in how CRAs are performed, using a well-known risk comparison—the allegation that exposure to the naturally occurring carcinogen aflatoxin was definitely and substantially riskier than exposure to the pesticide Alar—to demonstrate the implications of analytic overconfidence. From this example, general lessons will be gleaned to offer an improved paradigm for comparing environmental risks.

Background

CRA fell into some disrepute during the last decade, largely because one particular form of it, the quantitative contrasting of markedly dissimilar risks [such as being overweight versus being exposed to benzene (*1*)], was increasingly regarded as unresponsive to important perceptual judgments and hence as needlessly manipulative (*2,3*). Nevertheless, many other brands of CRA have flourished during the same period, while CRA of dissimilar risks seems to be making a comeback of late (*4,5*). In this regard, it is useful to distinguish between what could be termed "small" and "large" versions of CRA. The former involves the quantitative comparison of single risks that are generally less dissimilar than the overweight/benzene sort of comparison. Prominent examples of different types of "small" uses of CRA include the ranking of various hazardous waste sites in the Hazard Ranking System developed by the Environmental Protection Agency (EPA), the analysis of "risk/risk trade-offs" such as the choice between cancer risks due to the disinfection of drinking water and pathogenic risks due to the failure to disinfect (*6*), and the ranking of various common pollutants (both naturally-occurring and synthetic), either in order of inherent toxicologic potency or of excess risk under specified exposure conditions (*7*).

"Large" CRA involves the comparison of categories of risks and is increasingly being invoked as a means of putting the United States' allegedly haphazard environmental priorities in a "rational" sequence (*8–11*). For example, a recent magazine article cites as strong evidence that "we still haven't figured out what is really worth worrying about" the disparity between the $0.1 billion society spends annually on controlling indoor radon, which EPA estimates may cause as many as 20,000 lung cancer deaths each year, and the $6 billion spent on cleaning up hazardous waste sites, which purportedly cause fewer than 500 annual cancer deaths (*12*).

Misplaced Criticism of Comparative Risk Assessment

By far the most commonly criticized attribute of CRA is its alleged reliance on juxtaposing partially or totally incommensurable situations: the classic "you can't compare apples and oranges" problem [see Covello et al. (*13*) for this criticism applied to small CRA; see Hornstein (*14*) for an application to large CRA]. Although CRA is indeed difficult to do

and easy to botch, the major obstacle is not the qualitative differences between risks, but a completely different and largely ignored problem: the uncertainty in quantitative risk magnitude. Ironically, critics of CRA thus may well be right, but for the wrong reasons.

The impotence of the accusation of incommensurability is relatively easy to demonstrate. We all routinely compare highly dissimilar states by the simple (at least conceptually) cognitive process that involves: 1) disaggregating each situation or choice into its salient attributes (in the literal apple/orange comparison, these would be price, taste, nutritive value, appearance, etc.); 2) gauging how much of each attribute each situation or choice embodies; 3) assessing how much we value each attribute; and 4) aggregating the individual value judgments into a composite evaluation for comparison.

So apples and oranges are not incommensurable at all, and neither are disparate risks to health. In fact, when researchers tested empirically the most widely accepted predictions about how laypeople were supposed to react to various kinds of risk comparisons, the responses either did not support or else contradicted the thesis that the more dissimilar the comparison, the less acceptable and more aggravating the recipients would find it (*3,15*). For example, those surveyed by Roth et al. (*3*) generally regarded a hypothetical comparison of two different estimates of the same pollutant risk [a type of comparison Slovic (*15*) had put in their "first rank" of very acceptable communication techniques] as less reassuring, informative, and trust-engendering than a comparison of the pollutant risk with the risk of lightning, hurricanes, and insect bites (one of the Slovic et al.'s "fifth rank" or "rarely acceptable" comparisons).

The real problem in comparing risks is not that they differ in (known) qualitative respects, but that they differ in unknown quantitative respects. No amount of careful thought could make a choice between buying apples or buying oranges anything but arbitrary if one could neither discern nor control the price, taste, or appearance of either commodity. A numerical comparison between uncertain health risks, made without taking account of the uncertainty, is like shopping for produce sight unseen when one foodstuff might be expensive and rotten and the other cheap and flawless. And yet this is exactly how environmental risk assessors routinely make risk comparisons.

The further irony in this situation is that the analytic tool for making honest comparisons of uncertain risks—quantitative uncertainty analysis—is already well developed but languishes unused for this important application. For almost as long

as risk assessment has existed, researchers have used tools such as expert judgment, Bayesian analysis, and Monte Carlo simulation to estimate the uncertainty surrounding single risks (*16,17*). These uncertainties arise, among other sources, from our inability to measure precisely the quantities that drive the risk assessment models we use (parameter uncertainty) and from our inability to know which of two or more alternative models is in fact correct or most useful (model uncertainty). The most recent report on risk assessment by the National Research Council (*18*) contains numerous recommendations instructing EPA (which has lagged behind the advances in academia) to abandon its reliance on point estimates of risk for standard setting and to instead quantify uncertainty in risk using existing data and methodologies. However, none of the academic literature on uncertainty in risk, nor any of the practical applications conducted by EPA and other stakeholders in risk management policy, has ever applied the methodology to risk comparison.

This omission is particularly glaring because the mathematics of uncertainty dictate that dividing one uncertain risk by another to arrive at a comparative assessment magnifies rather than attenuates or cancels the uncertainty present in each risk (as long as the uncertainties do not arise from identical sources). For example, suppose you can guess the weight of person A to within a factor of 1.2 (e.g., your best guess is 180 pounds but you are confident A weighs between 150 and 216 pounds), and you can also guess the weight of B within a factor of 1.2 (e.g., your best guess is 150 and the range is between 125 and 180). Then, your best estimate of their relative weights would be 1.2 (180/150), but the uncertainty about this comparative estimate would range between 0.83 (150/180) and 1.73 (216/125). The uncertainty about the ratio estimate is now a factor of 1.44 on either side of the central estimate, larger than was present for either risk alone. Notice further that one cannot say with confidence that A weighs more than B. Thus, it is precisely for those applications where we can be least confident in our results that we devote the least effort to exploring how error-prone our answers might be.

Exploring Overconfidence in Risk Comparison

To develop and explore the implications of a more technically sound paradigm for CRA, I reexamined one of the most influential examples of small CRA: the conclusion reached by a group led by Ames (*19*) that the aflatoxin B_1 contained in a daily ration of peanut butter posed 18 times greater risk

than the growth regulator daminozide (Alar) in a daily ration of apple juice (a risk largely due to Alar's hydrolysis product unsymmetrical dimethylhydrazine, or UDMH, a potent rodent carcinogen). [This point estimate of risk has undergone some minor metamorphoses since it first appeared. Originally, Ames and Gold (*19*) presented the HERP (human exposure/rodent potency) index for aflatoxin (0.03%) as 17.6 times that of UDMH (0.0017%). Some weeks later, Ames cited a ratio of 10:1 (*20*), and later in 1989 then-FDA Commissioner Frank Young attributed to Ames a ratio of 30:1 (*21*). More recently, Uniroyal Chemical Company, the manufacturer of Alar, cited a ratio of 300:1 (*22*). In their most recent update of the HERP table (*23*), Ames and colleagues provided more information on the inputs to these numbers, but the implicit ratio remained essentially the same (0.03%/0.002%, or 15:1).] Whatever the precise number touted, it consists of the ratio of two risk estimates, each of which is composed of at least two uncertain inputs (at the highest level of aggregation, exposure and carcinogenic potency). Thus, any comparison of two HERP values (or other risk estimates) to generate a risk ratio entails calculating the uncertain quotient of two uncertain quotients. The sign and the magnitude of these estimates of the aflatoxin/Alar risk ratio have been cited to support the view that the "artificial" hazard of Alar is (or was) trivial compared to the magnitude of the risk from aflatoxin, a "natural" risk consumers supposedly deem acceptable (*24*).

It is conceivable, of course, that any estimate of this particular risk ratio, even if surrounded by a range of uncertainty, is meaningless because one or both of the substances involved are not carcinogenic in humans. A superficial look at Alar and aflatoxin might suggest that the latter is a "known" human carcinogen while the former is only known to cause tumors in rodents. But that would be a premature judgment. First, although in a few cases, such as saccharin and unleaded gasoline, directed research on chemical-specific mechanisms has cast serious doubt on whether certain animal carcinogens present any risk to humans at low doses, no such evidence or theory currently exists in the case of UDMH that would explain a qualitative interspecies difference. Besides, the lack of epidemiologic data (positive or negative) on UDMH does not necessarily distinguish it from an extensively studied chemical like aflatoxin. In no single case has "negative" epidemiologic data alone been of sufficient power to invalidate positive animal data (*25*); the fact that UDMH is not a "known" human carcinogen says more about what we know than about

what properties the chemical truly does or does not possess. In particular, the human data on these two substances may only differ because one (aflatoxin) is associated with a rare cancer (primary hepatocellular carcinoma) that stands out from the background, while the other may well increase the incidence of some more common tumor type(s) that could not be detected in a typical epidemiologic study. In any event, the method used here to quantify uncertainty in carcinogenic potency explicitly accounts for the additional uncertainty caused by the possibility that UDMH may pose zero or near-zero risks at low doses because we cannot be confident that the rodent tumors are relevant to humans. Finally, recently emerging evidence suggests that aflatoxin may not be a significant contributor to human liver cancer. Campbell et al. (26) claim that previous analyses of the epidemiologic data on aflatoxin were confounded by the failure to control for dietary variables and that aflatoxin is "an unnecessary and insufficient cause" as compared to viral and nutritional factors. The CRA presented here, like all previous ones, will not directly account for the model uncertainty contributed by the possibility that one or both contaminants are noncarcinogenic in humans, but will instead concentrate on the substantial amount of uncertainty present even assuming both substances pose non-zero risk.

Methods

The excess cancer risk to an individual consumer (i) of peanut butter or apple juice (j) is a function of three factors: 1) the amount of the foodstuff consumed each day (A_{ij}); 2) the concentration of aflatoxin or UDMH in the foodstuff (C_{ij}); and 3) the carcinogenic potency of each contaminant ($ß_j$). The first two of these quantities can be measured reasonably precisely, but they vary substantially among individuals; the third might be invariant across the population (if each person had equal biological susceptibility to the carcinogenic stimulus), but it clearly cannot be estimated without considerable ambiguity. With the appropriate units specified, risk is simply the product of these three quantities divided by the body weight of the individual (in this example, body weight was assumed to be invariant; the value 20 kg was chosen to represent a 4-year-old-child).

$$R_{ij} = [A_{ij} \text{ (g/day)} \times C_{ij} \text{ (ppb)} \times 10^{-6} \text{ (mg/ng)}] \times ß_j$$
(excess lifetime risk per mg/kg-day)]/20 kg

Point estimates such as the 18:1 risk ratio are derived by multiplying single values for consumption, concentration, and potency and reporting the quotient of the two resulting risk estimates as a single number. Since

each of the three inputs for each risk estimate can be described more correctly by a probability density function (PDF) than by an arbitrary point estimate, the raw material for a more sound approach to CRA entails first deriving these PDFs and then combining them to yield an estimate of the risk ratio with its associated uncertainty. Combining the PDFs is now computationally simple, with the advent of microcomputers to perform Monte Carlo simulation. In this method, a value from each PDF is chosen at random via an algorithm that ensures that the probability of selecting any value is the same as the underlying probability in the PDF. A single Monte Carlo iteration consists of a random draw from each PDF followed by the appropriate functional combination thereof (in this case, multiplication of three numbers to estimate each risk, followed by division of one risk estimate by the other). With repeated iterations (20,000 in this analysis), a PDF emerges for the output which asymptotically matches the distribution that would be obtained if the individual PDFs could be combined analytically [for this analysis, the Monte Carlo software "@RISK" (version 1.1 for Microsoft Excel, Palisade Corp., Newfield, New York) was used].

Data Sources

Food consumption. Data on the amount of peanut butter and apple juice consumed by children were obtained from a nationwide survey conducted by the U.S. Department of Agriculture (27). This survey of almost 38,000 persons, including 1,719 children ages 3–5, provides information on the average quantity of each foodstuff consumed each day, and also gives seven percentile points of the cumulative distribution of consumption across the population. In this analysis, the PDF for peanut butter consumption was well-approximated via a lognormal distribution with a median of 8 g/day and logarithmic standard deviation $\sigma_{ln} = 0.84$. The data on apple juice consumption were also well approximated by a lognormal PDF with a median of 83 g/day and a logarithmic standard deviation of 1.0. For reference, the point estimates of consumption Ames (19) apparently used (32 g/day for peanut butter and 120 g/day for apple juice) lie at approximately the 95th and the 64th percentiles of their respective PDFs. Without the distributional information, one would not be aware that these point estimates differ in their degree of "conservatism" (in such a way as to help make aflatoxin seem riskier than Alar), or that neither estimate reasonably approximates the amount of each food eaten either by frequent or by sporadic consumers.

Residue levels. Data on aflatoxin levels in 44,788 samples of peanut butter made

from the 1986, 1987, and 1988 peanut crops were provided by the National Peanut Council (28). Data from the three crop years were combined to yield a discrete distribution consisting of 13 different possible residue levels and their associated probabilities; the overall mean of this distribution was 2.82 ppb (this distribution was approximately lognormal in shape, but because it had a slightly shorter right-hand "tail" than the continuous distribution would have yielded, the measured discretized values were used instead). The point estimate of concentration used by Ames [2 ppb (19)] lies at approximately the 40th percentile of this distribution. In contrast, *Consumer Reports* noted in 1990 (29) that 86 samples of peanut butter tested averaged 5.7 ppb of aflatoxin. However, they deliberately oversampled from less well-known brands (30).

Residue levels for UDMH in apple juice were provided courtesy of the Uniroyal Chemical Company (31). Uniroyal analyzed 71 samples of apple juice for UDMH content; the juice came from the 1985 or 1986 apple crops. The sample mean was 13.8 ppb, and the maximum concentration was 83 ppb. [There is a separate category of "baby apple juice," the small jars that infants (and some toddlers) consume. The mean UDMH content in the 71 samples of baby apple juice was nearly twice that of the adult product, and the maximum single value was 112 ppb (31). Thus, using only "adult" apple juice data tends to underestimate both the relative and absolute risk of UDMH exposure.] Due to the small number of samples and the fact that the data clumped into at least four modal groups (35 of the 71 values were clustered either around 1, 8, 13, or 33 ppb), the PDF used in the analysis consisted of the data points themselves; in the Monte Carlo procedure, 1 of these 71 values was chosen at random at each iteration. Ames (19) apparently assumed that apple juice always contains about 7.5 ppb UDMH; this value lies at about the 45th percentile of the distribution of measured residue levels.

Carcinogenic potency. The most difficult portion of the analysis was the generation of the PDFs for cancer potency, as no standard methods currently exist for deriving such distributions (32). Two different methods were used here, reflecting the distinction between a "known human carcinogen" (aflatoxin) and a substance (UDMH) for which only animal bioassay data are available.

The distribution for the potency of aflatoxin (Table 1) was derived from a risk assessment recently completed by the California Environmental Protection Agency (CalEPA) (33), which made use of

new epidemiologic data compiled by Yeh et al. (*34*). This was a cohort study of approximately 8000 persons in Guangxi, China, examining the relationship between aflatoxin exposure and primary hepatocellular carcinoma, controlling for concurrent infection with the hepatitis B virus (HBV). CalEPA tested five mathematical models and recommended the interactive effects form of the excess risk model, based on its fit to the Guangxi data, the stability of the parameter estimates obtained, and its ability to predict liver cancer incidence in the United States given reasonable assumptions about HBV prevalence and aflatoxin exposures. [The interactive excess risk model has the form $y = a + \beta_1 H + \beta_2 d + \beta_3 Hd$, where y is liver cancer incidence, a is the background incidence (in the absence of HBV infection or aflatoxin exposure), d is the daily dose of aflatoxin, H is a dummy variable indexing HBV carrier status (1 = positive, 0 = negative), and the β_i are fitted coefficients representing the HBV effect, the potency of aflatoxin, and the interactive effect, respectively.] Using the CalEPA regression equations and the standard errors they reported, maximum likelihood estimates (MLEs) and 5th and 95th percentile values for β^- (the potency of aflatoxin in an HBV-negative person) and β^+ (the potency in an HBV-positive person) were derived (see Table 1). In the @RISK spreadsheet, these normal distributions were truncated at zero so that negative values for potency could not occur. At each iteration in the Monte Carlo simulation, the potency of aflatoxin is determined with reference to f, the assumed prevalence of HBV-positive individuals in the population. According to the CalEPA report (*33*), plausible values for f in the U.S. population range between 0.1% and 1%; a value of 1% for f was chosen here, an assumption that tends to overstate the relative and absolute risk of aflatoxin exposure. The Monte Carlo process then randomly chooses values from either the β^+ or β^- PDFs, in a 1:99 ratio, thereby preserving the bimodality of the PDF for the potency of aflatoxin to a randomly chosen person in the population.

The PDF for the potency of UDMH is derived by a rather different procedure because no human data exist for this substance. There are various troublesome sources of uncertainty in analyzing an animal bioassay and extrapolating the results to humans, including the choice of dose–response model, interspecies scaling of exposure and susceptibility, and random sampling error affecting the small groups of rodents tested. EPA pays some attention to the last of these three uncertainties by publishing the 95th percentile upper confidence limit (UCL) on the slope of the lin-

earized dose–response function that fits the observed tumorigenicity data acceptably well. In addition, EPA usually includes the caveat that the true slope at low doses "could be as low as zero." There are several problems with this approach: 1) for each case, it provides the risk manager and the public no idea how likely the UCL, zero, and all values in between are to be true, or even whether the value zero is plausible at all; 2) it gives no information on the nature and implications of the 5% of the distribution above the UCL; 3) it does not allow for nonlinear dose–response functions, in effect treating "potency" as a scalar independent of dose; and 4) it assumes, probably incorrectly, that the asymptotic confidence limits (derived by examining changes in the log-likelihood function with reference to the χ^2 distribution) are valid for the case of small samples and constrained (non-negative) optimization of the regression coefficients (*35,36*).

I have adapted work of Guess et al. (*35*), Sielken (*37*), and others to develop a method for quantifying potency uncertainty that addresses these four problems [but that, like EPA's approach, does not deal fully with model uncertainty in dose response (e.g., the possibility that a threshold exists) or in interspecies scaling (*38*)]. The method involves performing a bootstrap analysis of the observed bioassay data. For example, if the original bioassay had a single positive dose group in which 20 animals out of 50 tested developed tumors, the simulated bioassays would have tumor responses ranging from perhaps 15 to 25 animals, depending on the assumption made about the sampling error inherent in the single data point. If 10,000 such simulations were generated, and the resulting

Table 1. Input probability density functions for potency of aflatoxin[a]

	HBV status	
	Negative	Positive
Mean	15.4	202.9
SD	7.3	55.7
5th Percentile	3.4	111.3
95th Percentile	27.4	294.5

[a]Potency is expressed in units of $(mg/kg\text{-}day)^{-1}$; the values in this table have been standardized via the surface area correction to apply to a 20-kg child.

(linear) dose–response functions were put in ascending order of steepness, the 500th highest observation of the slope of the line would provide an alternative estimate of the 95th UCL of potency. The method uses the computer program "MSTAGE87" (version 1.1, courtesy E. Crouch, Cambridge, Massachusetts) to calculate the best-fitting polynomial for each simulated data set. By keeping track of all the coefficients, potency can depend on higher-order terms when the linear term is estimated to be near zero (i.e., the distinction between "the potency is zero" and "the dose–response curve is quadratic at low doses" is not muddled).

The bootstrap uncertainty analysis was applied to a new bioassay of UDMH carcinogenicity sponsored by Uniroyal (*39*). Table 2 shows the results of the new UDMH bioassay; because individually coded data for each test animal were not available, only the primary tumor response (hemangiosarcomas plus hemangiomas) was considered, not the total number of animals with tumors at any site (this would include pulmonary neoplasms as well). CalEPA recently completed an analysis of this bioassay (*40*) and calculated a potency value somewhat higher than EPA's. CalEPA used the tumor site that gave the highest UCL for potency, namely, pulmonary carcinomas/adenomas; here, EPA's assumptions about the appropriate data set were used, largely because the blood vessel tumors were so rare in the control animals, in contrast to the pulmonary tumors (35/100 pulmonary tumors among controls, as opposed to only 5/100 vascular tumors among controls).

The bootstrap resampling consisted of 5000 simulated data sets (see Table 3). The fitted values for β_1, the linear term in the multistage polynomial, ranged from 0 (5.2% of all values) to 1.54; the median value for β_1 was 0.508, and the 5th and 95th percentiles were 0 and 0.850, respectively. The PDF is approximately normal, as would be expected when the observed bioassay data are roughly linear; when the observed data can best be fit by a polynomial with no linear term, the PDF for β_1 is approximately exponential in shape (*38*). The 5000 pairs of β_1 and β_2 values were sampled at random in the Monte Carlo process. The risk of

Table 2. Data on carcinogenicity bioassay of unsymmetrical dimethylhydrazine (UMDH) in male mice (*39*)

UDMH concentration in drinking water (ppm)	Dose (mg/kg-day)	Human equivalent dose (mg/kg-day)[a]	Incidence of hemangiosarcomas/ hemangiomas
0	0	0	5/66
40	7.34	0.8797	31/67
80	13.7	1.65	43/68

[a]Rodent doses were converted to the equivalent doses for a 20-kg child by dividing by the factor $(20/0.035)^{1/3}$, where 0.035 kg is the average weight of a male mouse. This procedure assumes that children and mice are equally susceptible to UDMH on a dose per surface-area basis.

UDMH exposure was calculated at each iteration as $(\text{\ss}_1 d + \text{\ss}_2 d^2)$, where the dose d was defined as intake × residue concentration/body weight). Thus, the computations account for the possible sublinearity of the UDMH dose–response function and permit some probability that the risk to humans at low doses is essentially zero.

Note that the new bioassay data give similar values for the potency of UDMH to those of the controversial Toth study (*41*), although the extent to which the new study should be interpreted as confirming, modifying, or invalidating the earlier one still seems to be a subject of controversy (*42–44*). The maximum likelihood estimate and UCL for $\text{\ss}_1$ in the Toth study (if adjusted to a 20-kg child) were 0.680 and 0.907, respectively.

Table 4 summarizes some key parameters for each of the six input distributions. The sizes of the uncertainties in these parameters are typical of those encountered in previous assessments of the uncertainty in risks assessed singly. Several of the parame-

ters have rather "tight" distributions (i.e., their 95th-percentile values are less than 10 times higher than their 5th percentile values), while one (UDMH residue) varies by nearly 100-fold, and another (UDMH potency) is "infinitely uncertain" in the sense that its lower bound could be zero. For comparison, Finley et al. (*45*) suggested distributions for 12 of the parameters commonly encountered in more complicated multimedia exposure assessments. Some of the distributions they recommend are as tight as some of those in Table 4 (e.g., inhalation rates among adults vary between approximately 8 and 16 m³/day, to a 90% degree of confidence), while others (e.g., the number of years an individual is likely to live at one residence before moving) vary by more than 100-fold, and still others (e.g., the amount of soil a child ingests each day) resemble the UDMH potency distrib-

ution in that there is a nontrivial probability that zero is the true value.

Results

Figure 1 shows the cumulative probability distribution functions (CDFs) for the excess lifetime risks of peanut butter and apple juice consumption. Selected summary statistics of these distributions are presented in Table 5. The CDF for UDMH risk has a slightly higher median than the aflatoxin risk CDF, but because the former distribution has a much longer right-hand tail, its mean is nearly twice as high as the latter distribution's mean value.

The effect of this overlapping of the two risk distributions is shown in Figure 2, which depicts the PDF for the common logarithm of the ratio of the UDMH risk to the aflatoxin risk. Several features of this PDF are noteworthy, in light of the deter-

Table 3. Summary statistics for uncertainty in carcinogenic potency ($\text{\ss}_1$) of unsymmetrical dimethylhydrazine (UMDH) (adjusted to 20-kg child)

	CalEPA data set	U.S. EPA data set
Probability $\text{\ss}_1 = 0$	3.9%	5.2%
5th Percentile	0.067	0
10th Percentile	0.26	0.13
25th Percentile	0.60	0.35
50th Percentile	0.82	0.51
75th Percentile	1.0	0.62
90th Percentile	1.18	0.73
95th Percentile	1.35	0.85
Asymptotic[a] (5th, MLE, 95th)	0.36,0.92,1.16	0.20,0.58, 0.72

[a]The asymptotic values for $\text{\ss}_1$ were calculated in the same manner as EPA does, by determining the slope of the linearized dose–reponse function that maximizes the likelihood function given the observed bioassay data (the maximum likelihood estimate; MLE), and then increasing or decreasing the linear coefficient of the dose-response function until it could be rejected as not fitting the data at an upper or lower $p = 0.05$ level of confidence (via reference to the χ^2 distribution). Note that the bootstrap resampling technique described in the main text yields distributions that are somewhat broader than those generated by the EPA method.
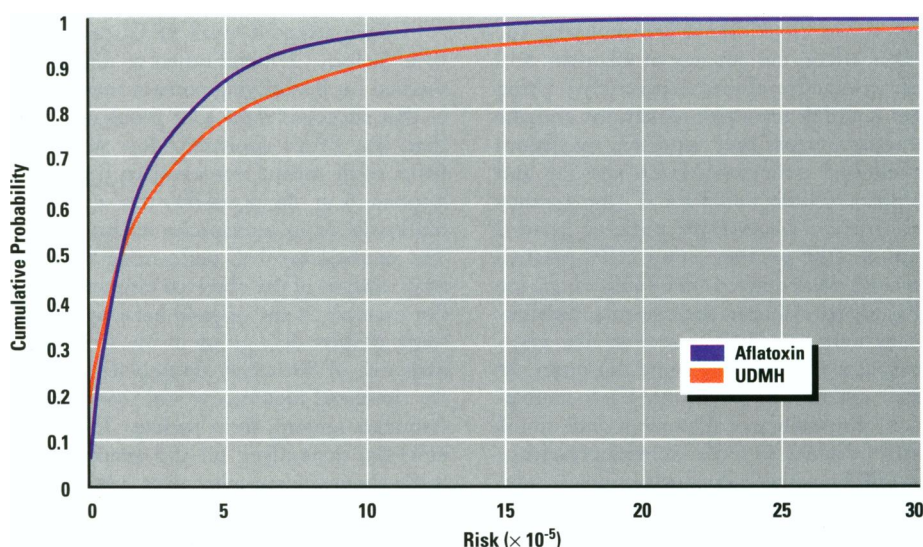
Figure 1. Cumulative distribution functions (CDFs) for the excess lifetime risk of peanut butter (blue curve) and apple juice ingestion (red curve). In either curve, the X-coordinate corresponding to a given value on the Y-axis represents the risk level that with probability y is less than or equal to the true but unknown value of risk. For example, the curves intersect at approximately $y = 0.5$, so there is roughly a 50% chance that either risk is less than about 1.3×10^{-5} (see Table 5 for a tabular representation of this figure). The red curve lies below the blue curve above $y = 0.5$, which means that as one approaches "worst-case" conditions, unsymmetrical dimethylhydrazine (UDMH) is (much) riskier than aflatoxin (e.g., there is a 5% chance the risk of aflatoxin exceeds 1×10^{-4}, whereas continuing horizontally from $y = 0.95$, the UDMH curve is not intersected until the risk level equals 2×10^{-4}).

Table 4. Characteristics of the probability density functions (PDFs) for the input variables

Variable	Units	Median	Mean	5th %ile	95th %ile	Uncertainty factor[a]	Percentile location of mean[b]
Peanut butter consumption	g/day	8.00	11.38	2.00	31.86	15.93	66
Apple juice consumption	g/day	83.00	136.84	16.02	430.02	26.84	69
Aflatoxin residue	ppb	2.50	2.82	1.00	6.50	6.50	61
UDMH residue	ppb	9.00	13.75	0.5	42.00	84.00	67
Aflatoxin potency (population average)	(mg/kg-day)$^{-1}$	15.40	17.50	4.02	28.23	7.02	61
UDMH potency (linear term only)	(mg/kg-day)$^{-1}$	0.508	0.490	0	0.850	—	43

UMDH, unsymmetrical dimethylhydrazine.
[a]Ratio of the 95th to the 5th percentiles.
[b]Percentile of the PDF where the arithmetic mean is located; a measure both of skewness and heaviness of tail.

ministic point estimates of Ames and others that this ratio is approximately 1:18.

The central tendency estimates (both the median and the mode) of this ratio are virtually indistinguishable from 1:1. This indicates a comparative risk for apple juice consumption at least an order of magnitude higher than any of the point estimates cited (*19–23*). Contrasting this result with previous risk comparisons reveals another intrinsic flaw in the use of point estimates. Because previous investigations failed to place the point estimates of inputs and results in context (i.e., were they central, lower-bound, or upper-bound numbers?), it is unclear whether the difference between 18:1 and 1:1 is due to a shift in the conservativeness of these estimators, due to changes in the input data (e.g., the newer bioassay of UDMH), or both.

More important than any single estimate of the comparative risk is the large uncertainty revealed here to affect that comparison. It happens that the central estimate of this particular risk ratio is so close to unity that it is clearly reckless to conclude that either risk is definitely greater than the other. The faint signal that aflatoxin may on average be 1.03 times riskier than UDMH is far outweighed by the "noise" in the comparison, which extends over four orders of magnitude at a 90% confidence level (from 376:1 in favor of aflatoxin to 34:1 in favor of Alar, a difference of a factor of 12,700). A nonparametric measure of the amount of overlap in the two risks was also computed to supplement this comparison of the median of the risk ratio to its own variance. By the Wilcoxon rank-sum test (*46*), the two risk PDFs are indistinguish-

able ($z = 0.525$, $p \approx 0.3$), so the hypothesis that the two PDFs are different must be rejected. Readers who sense that there is a paradox here (how can the two risks be simultaneously "the same" and yet differ by 30 or 300-fold?) may be caught in a semantic trap. There is no inconsistency in believing both parts of that statement. It is the distributions that are statistically indistinguishable; since the true value of either risk could fall anywhere within its own PDF, two independent risks with similar PDFs may, in fact, differ wildly.

The major point of this article (and of improving CRA in general) is not to engage in "dueling point estimates," but to progress beyond any single point estimate comparison by changing the currency with which risks are expressed. In other words, this analysis shows that 18:1, 1:1, 1:18, and other answers are all legitimate, but that none of them alone expresses the risk correctly. Assuming this analysis is computationally sound, the only informative way to express the comparative risk of aflatoxin and Alar is to acknowledge the multiplicity of legitimate quantitative conclusions. A statement such as "to a reasonably high degree of confidence, aflatoxin is no more than 376 times riskier than Alar; on the other hand, Alar could be as much as 34 times riskier than aflatoxin" (see Table 5) has the virtue of candor and of revealing the complexity of any decision to control (or be concerned about) one or the other substance preferentially. Its drawback, that it does not lend itself to black-and-white conclusions, is equally prominent, but one must balance the tidiness of a point estimate such as 18:1 against the virtual certainty that

other comparative risk estimates (and hence other social or personal decisions) are at least equally valid.

The impact of the uncertainty on the comparative risk assessment is robust to computational differences between this and previous analyses and to assumptions about the human carcinogenicity of UDMH. Again, even though the contrast between 18:1 and 1:1 is subsidiary to the larger difference between point estimates and expressions acknowledging uncertainty, Figure 2 reveals that even if this analysis suffered from a hidden systematic flaw that biased it toward overstating the relative risk of UDMH (which I argue is not a strong possibility), the general point still stands that a facile comparison is vulnerable to serious error. Suppose, for the sake of argument, that such a hidden flaw was found and the entire PDF in Figure 2 was shifted 18-fold to the left (that is, matching the central tendency exactly to Ames's 18:1 estimate). There would still be a roughly 10% chance that apple juice was riskier than peanut butter, and a roughly 1% chance it was more than 10 times riskier. Similarly, even if those convinced that UDMH is not a human carcinogen (see above) could successfully argue (presumably bringing to the table some concrete evidence, either direct or indirect) that there was a 90% probability its risk was zero, there would still be about a 5% chance that UDMH was riskier than aflatoxin. It is entirely a question of policy and values, not of science, whether even an analysis that might have shown such a 90/10 or 95/5 split could legitimately be reduced to the overconfident pronouncement that "peanut butter is riskier."

The PDF is not obviously biased toward overstating or understating the extent of uncertainty. Since the three factors analyzed are only some of the major uncertainties and variabilities affecting these two risk assessments, the results presented here might well understate the true ambiguity in the risk ratio. For example, the analysis assumes that every person is equally susceptible to the carcinogenic effects of aflatoxin or of UDMH; this assumption, though commonly made, ignores evidence that inborn and acquired variations in enzymatic metabolism, DNA repair, immune surveillance, and other factors cause individual susceptibility to cancer for a given exposure to vary widely [with perhaps three to four orders of magnitude separating the most susceptible and least susceptible portions of the "normal" population (*47*)]. This omission tends to bias both of the absolute risk estimates downward (*18*). On the other hand, there are several features of this analysis that

Table 5. Summary statistics for output distributions ($N = 20,000$)

	UDMH risk	Aflatoxin risk	Ratio UDMH/aflatoxin[a]
Minimum	0	3.99×10^{-9}	0
1st percentile	1.97×10^{-11}	4.42×10^{-7}	1.58×10^{-6} (1:632,911)
2.5th Percentile	4.14×10^{-10}	9.26×10^{-7}	3.43×10^{-5} (1:29,155)
5th Percentile	6.38×10^{-8}	1.46×10^{-6}	2.66×10^{-3} (1:376)
10th Percentile	6.15×10^{-7}	2.35×10^{-6}	2.79×10^{-2} (1:36)
25th Percentile	3.22×10^{-6}	5.31×10^{-6}	0.193 (1:5)
50th Percentile	1.33×10^{-5}	1.26×10^{-5}	0.972 (1:1.03)
75th Percentile	4.29×10^{-5}	2.95×10^{-5}	4.33
90th Percentile	1.09×10^{-4}	6.18×10^{-5}	15.96
95th Percentile	1.83×10^{-4}	9.57×10^{-5}	33.88
97.5th Percentile	2.97×10^{-4}	1.39×10^{-4}	64.52
99th Percentile	4.81×10^{-4}	2.15×10^{-4}	139.82
Maximum	5.97×10^{-3}	1.44×10^{-3}	17,545.6
Mean[b]	4.60×10^{-5}	2.72×10^{-5}	

UMDH, unsymmetrical dimethylhydrazine

[a]Note that the values in the third column of numbers are not simply the quotients of the numbers in the first and second columns; the third column contains the summary statistics of a separate probability density function (PDF) derived from the Monte Carlo simulation that takes into account the possibility that one risk truly lies in the left-hand tail of its own PDF while the other true value lies in its own right-hand tail (and, with equivalent probability, vice versa).

[b]Note that the arithmetic mean of the distribution of ratios is a nonsensical statistic, and hence is not reported here. Since ratios are inherently geometric (as opposed to arithmetic) quantities, their arithmetic mean gives disproportionate weight to cases where the numerator exceeds the denominator, and is thus entirely an artifact of which risk is placed in the numerator (i.e., the means of A/B and of B/A might both be greater than 1).
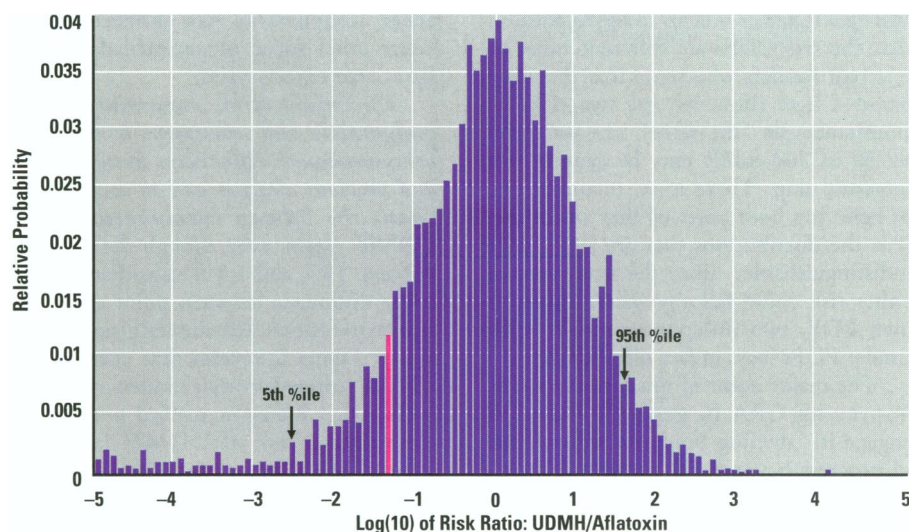
Figure 2. Probability density function (PDF) for the ratio of the risk of unsymmetrical dimethylhydrazine (UDMH) to the risk of aflatoxin, generated via 20,000 realizations from the Monte Carlo simulation described in the text. The X-axis denotes the common logarithm of the risk ratio (hence, $x = 3$ represents a risk of UDMH 1,000 times that of aflatoxin; $x = -2$ represents a risk ratio of 100:1 in the opposite direction). The height of the histogram at any point denotes the relative probability of that value compared to other possible values (the area under the smooth curve approximated by this histogram equals unity). The deterministic point estimate (1:18) of Ames et al. (*19*) lies at $x = -1.25$ (pink bar); the 5th and 95th percentiles (as shown by the two arrows) lie at $x = -2.57$ and $x = 1.53$, respectively.

might contribute to an overstatement of the total uncertainty in the comparison. Notably, the use of the entire distribution of measured residue data implicitly assumes that some consumers ingest products with high (or low) contaminant levels day after day, rather than being exposed at random to the whole spectrum of contaminant concentrations over long periods. Because residue levels are correlated to some degree with brand name and with geographic market, this assumption may not be far off the mark. Similarly, the ingestion rates for peanut butter and apple juice may not be statistically independent. If avid consumers of one product tend to be high consumers of the other (the more peanut butter ingested, the more liquid needed to wash it down?), this analysis would overstate the variability in the ratio of the two exposures.

Using alternative assumptions or data sets could also shift the entire PDF upwards or downwards (without affecting its variance). For example, the central estimate of the UDMH/aflatoxin ratio would increase by approximately a factor of two if the tumor site chosen by CalEPA (*40*) was used to analyze the UDMH bioassay data, and it would have increased further if more recent data reflecting increased apple juice consumption in the United States during the 1980s had been available or if residue levels in "apple juice for infants" had been included (*31*). Similarly, it is plausible that the estimate of UDMH's potency, like all current estimates based on rodent studies, is approximately seven times

lower than the true value because the animals were only exposed for less than 2 years out of their natural life span (*48*). However, the fact that other researchers could legitimately generate other PDFs for the risk ratio, differing in mean and/or variance from the one in Figure 2, is not a weakness of this analysis. Rather, it illustrates a fundamental point: arguments about the precise extent of uncertainty reveal the bankruptcy of the practice of expressing risks and risk ratios via point estimates that admit no possible imprecision. The type of analysis undertaken here differs in kind as well as result from previous analyses: disputes about which point estimates are "correct" are resolved by the improved type of analysis. Remaining disagreements about exactly how to compute the uncertainty are not trivial, but they are second-order questions that should not obscure the fact that *a priori*, any CRA that results in a distribution of values is superior to any one that yields only a single value. To argue otherwise is to claim that scientists, decision-makers, and the public are better off with a guess that hides the fact it is a guess (i.e., the point estimate of relative or absolute risk) than with an analysis that attempts, but may conceivably fail, to precisely quantify the magnitude of the guesswork.

Discussion

These results have some important implications for risk management and risk communication, both specific to the case analyzed and for the general goal of improving decisions based on CRA.

Both Alar-contaminated apple juice and aflatoxin-contaminated peanut butter posed small, but not insignificant, individual and population risks. A central-tendency estimate of either risk falls at approximately 10^{-5}, an order of magnitude higher than the one-in-one-million benchmark generally regarded as the dividing line between minimal and significant risk when the exposed population is large. However, while these central tendency estimates are only mildly troubling, the UCLs lie in the range of 1 in 10,000 to 1 in 5,000. Thus, for the deliberately added contaminant (Alar/UDMH), a substantial number of preventable deaths may have been occurring each year in the population of children consuming apple juice, and a substantial number of children may have been subjected to risk levels several hundred times higher than those generally deemed to be acceptable. [For a rough calculation, consider only the subset of the approximately 2.5 million children born each year in the United States who faced risks from UDMH of at least 2×10^{-4}/lifetime. If this group made up 5% of each cohort, as Table 5 suggests, then each year there would be at least 25 excess deaths attributable to UDMH exposure in this subgroup. This figure is roughly numerically equivalent to the annual number of children murdered in public schools (*4*), a problem most people view as quite serious.]

Ironically, despite the various differences in the underlying data and despite the fact that this is a quantitative uncertainty analysis rather than a point-estimation exercise, the UCL in Table 5 is very close to the "plausible upper bound" of 1 in 4,000 that the Natural Resources Defense Council computed for UDMH in its much-maligned "Intolerable Risk" report (*49*). Actually, neither a Monte Carlo simulation nor a complicated point-estimation exercise is necessary to derive the approximate 1 in 4,000 lifetime risk estimate. Simply multiplying estimates for consumption (two 8-oz glasses per day) and for residue level (20 ppb) that each lie between the mean and the reasonable upper bound of their distributions yields a dose estimate about 2,000 times smaller than the surface-area-adjusted dose (approximately 1 mg/kg/day) that produced about a 50% tumor incidence in mice (in two different studies). As long as the assumption of proportionality is reasonable, 1/2,000 of this TD_{50} represents a risk of (0.5) (1/2000), or approximately 1 in 4,000 (*44*).

Overconfident pronouncements are not a good way to rank risks for decision-makers and the public. A social/political controversy continues to ferment as to whether numerical comparisons of risk should be used in isolation to inform peo-

ple what to worry about (*11*), given that people may legitimately regard a "smaller" risk as more worthy of attention or control than a "larger" one, depending on factors outside the purview of such quantitative rankings (e.g., issues involving dread, feasibility of control, locus of responsibility, and distributional equity). Even if analysts could somehow be sure that their numerical results would be used to supplement rather than monopolize this much larger priority-setting arena, however, they remain responsible for at least reporting in a thorough and honest fashion the narrower comparisons they purport to make. In this case, even if all other relevant factors had no effect on the risk comparison, it would be misleading to declare peanut butter the larger risk, when there is a 50% chance (if this analysis is exactly correct) or a 10% chance (even if this analysis is off by 18-fold in a particular direction) that such a statement is not true, even in a limited numerical sense. Recently, Ames' colleague Gold has claimed that their body of work on risk comparison was not designed to make or to encourage quantitative risk comparisons (*50*). Gold states that because of their well-known belief that there is little or no scientific basis for extrapolating from animal bioassays to human environmental risks, readers of their papers understand that they are not actually presenting risk estimates, but "merely ranking possible hazards." If these rankings are so uncertain as to be meaningless, however, then why express all the HERP indices to two significant figures, and why write that "the public might be better served if EPA were to present its risk assessments as comparisons to its estimates of risks from cups of coffee, beers, and so forth" (*51*)? A number cannot simultaneously be both extremely precise and infinitely uncertain; I maintain that quantitative uncertainty analysis is far superior to point estimation, no matter how many retrospective caveats are later placed on the point estimates.

The problems created by overconfident point estimates only increase with large CRA, because the kind of risks EPA, the U.S. Office of Management and Budget, Congress, and others wish to rank are much less straightforward to compare than even this rather uncertain comparison of two carcinogenic food contaminants. Returning to the radon/hazardous waste example cited at the beginning of this article, the CRA presented here should cast doubt on the definitive statements from EPA and the media that radon is exactly 40 times "worse" (or 2,400 times less efficient, if the additional and uncertain dimension of cost is also included) than the Superfund problem.

Uncertain risk comparisons, despite their complexity, are much preferable to avoiding quantification altogether. Although it is always simpler to criticize a misleading practice than to thoroughly describe a practical alternative, there are three cornerstones of decision-making under uncertainty that should help improve the way we calculate and communicate environmental and health risk comparisons. The message of this article is certainly not that we should eschew priority-setting—that would itself be contradictory, as priorities set by default or inertia are no less real than ones set consciously. Rather, the goal is to understand how formal analysis can inform priority-setting and where it must leave off and allow for creativity and subjectivity.

First, individual and social decision-makers must use the depiction of uncertainty to evaluate the probabilities and the consequences of making errors in their decisions, not just as another tool to answer an intellectual question about the magnitude of two disembodied problems. The decision determines how confident one needs to be that the larger risk is indeed larger. If the stakes are not high and large errors are not extremely more dangerous than smaller errors, then the central tendency of the risk ratio may be enough to go on. For example, if you are most concerned about picking the fruit with fewer calories, it may be sufficient to know that the average apple has, say, 80 calories to the average orange's 90, even if both values can range 30 calories above or below their averages. In this hypothetical case, you might be content to be only reasonably sure that apples were less caloric than oranges, given that even the worst portion of the rest of the distribution (the apple really has 110 calories to the orange's 60) does not lead to a decision costly enough to outweigh the benefits of being right on average. On the other hand, high stakes and/or asymmetries in the decision problem make it more important for the thoughtful decision-maker or risk communicator to consider the full range of possibilities and carefully evaluate which decision is best, rather than simply which risk is larger.

For the practical rather than the intellectual exercise, risk management thus involves, among other goals, trying to minimize the regret associated with the chosen option (where "regret" is a personal judgment related to the various costs incurred if the option chosen turns out to be inferior to another available one) (*52*). In the Alar/aflatoxin example, the question of which risk is worse is only a proxy for the real question of what to do about either or both substances. In the latter context, and given the results in Table 5, the individual

or the regulator must balance, say, the 5% chance that ignoring or delaying action on Alar would erroneously leave unaddressed a problem 34 times greater than aflatoxin, against an equal chance that the opposite decision would focus attention on a problem 376 times smaller than aflatoxin (or, assuming that nothing more can be done about the natural carcinogen, the choice becomes one between some probability of spending resources on a problem manyfold smaller than a background risk already accepted by society, versus ignoring a problem erroneously deemed smaller than the tolerated risk). Again, needlessly definitive statements that one of these risks is exactly x times worse than the other robs the listener of the knowledge that the simplistic choice might be wrong by any criterion he might use to value the risks.

Second, decision-makers and analysts also need to understand that there is nothing wrong with using point estimates to inform and simplify their tasks. After all, the quantitative aspect of environmental decisions hinges on numbers, not on abstract curves that subsume an infinite set of discrete estimates. But different kinds of point estimates are appropriate for different decision-making goals, and the unwitting choice of an estimate can confound the decision. For example, if the decision-maker's goal here was simply to maximize the probability of addressing the larger risk, then the median of the risk-ratio PDF would be the appropriate anchor, and either of the possible decisions would have a virtually identical error rate. If the goal instead was to minimize the expected cost of the decision (assuming cost was proportional to the true absolute difference between the two risks, so that incorrectly ignoring a much larger risk would be costlier than ignoring a slightly larger risk), then a comparison of the means of each PDF would be appropriate, and Alar would emerge as the higher priority. And if the goal was to minimize the chance of an extremely bad decision, the appropriate choice of a summary point estimate would depend on whether the decision-maker was more averse to gross errors of overspending or underprotecting (or to errors that favor ignoring a deliberately added contaminant versus those that favor ignoring a naturally occurring toxin). Because the UDMH risk distribution has both a longer right-hand tail and a longer left-hand tail than the aflatoxin PDF, either risk could be the priority depending on which percentile (near the 5th or near the 95th) corresponded to the eventuality the decision-maker particularly wished to avoid.

Finally, optimal decision-making requires careful attention to the twin influences of uncertainty and interindividual

variability. The latter is a property of the system being studied which causes different estimates to be valid for different individuals (and which is generally irreducible through further study); the former is a property of the investigator (and his limited knowledge of the system) which generally can be reduced through further study (*18*). The results presented to this point deliberately intermingle uncertainty and variability. For societal decision-making, the two phenomena can be usefully combined. The PDFs in Table 5 essentially represent the uncertainty in risk to a person selected at random from the exposed population. Thus, the fact that the 95th percentile risk estimate for UDMH is 1.83×10^{-4} does not necessarily mean that 5% of the population faced risks at this level or higher, nor does it necessarily mean there was a 5% chance everyone's risk was this large; rather, it means that knowing nothing about the consumption habits or exposure history of an individual, there is a 5% chance his or her individual risk was above this value. Similarly, the mean of 4.6×10^{-5} can be interpreted as $1/N$ times the expected number of excess deaths in a random population of N persons exposed to UDMH.

The PDFs summarized in Table 5 are really made up of a family of uncertainty distributions, which average out to the composite statistics presented; each distribution is applicable to a person at a particular fractile of the underlying variability distribution. For example, one could replace two of the three input PDFs in the spreadsheet (for consumption and residue levels) with deterministic values and arrive at statements of the following type: for an individual whose exposure to UDMH puts him at the 95th percentile of the population, there is an 80% chance (here due entirely to uncertainty in carcinogenic potency) that his risk is between 1.1×10^{-4} and 6.5×10^{-4}, with a median value of 4.4×10^{-4}. Therefore, even though both variability and uncertainty are irreducible (if a decision must be made today) from the government's vantage point, the individual can reduce uncertainty by considering where he or she falls in the population with respect to the characteristics that are variable. Of course, some of the components of variability in this example are easier to resolve than others. Although it would not be apparent from the definitive nature of the 18:1 pronouncements, a frequent peanut butter consumer might realize that in relative terms, Alar was even less of a problem than this assessment suggests, and conversely for the frequent apple juice consumer (whose absolute risk might closely approximate the narrower PDF referenced earlier in this paragraph). Even an individual's relative risk due to residue lev-

els might to some extent be clarified, as government or private organizations could analyze and publish the variation in residue levels by region, brand, or type of product (e.g., store-bought peanut butter versus the more highly contaminated "grind-your-own") (*29*). And, at least in the case of aflatoxin risk, motivated citizens could learn more about their own biologic susceptibility (to the extent that tests for hepatitis B virus antibodies accurately indicate higher risk).

Social decision-makers can also profit from attempts to decouple uncertainty and variability, as they can then intelligently rephrase the questions at hand. The questions of whether aflatoxin is riskier than Alar or whether radon is a bigger problem than Superfund sites are needlessly overaggregated; both for thorough risk communication and for more creative control strategies, more useful questions would be, for whom is risk A worse than risk B? Thus, rather than declaring that radon abatement should increase at the expense of waste-site cleanups, EPA might try to identify particular situations where marginal decreases in risk from the latter problem might be foregone to target efforts at "hot spots" of radon risk. Similarly, in situations where societal decision-makers wished to compare risks solely based on their expected population consequences (i.e., without regard to individual risk levels or their distribution), substituting deterministic average values for consumption and residue levels would yield a narrower distribution measuring only the uncertainty in the expected number of excess fatalities. In this case, of course, the aflatoxin and UDMH distributions would still substantially overlap.

Conclusions

CRA will never be both technically valid and acceptable to citizens and government unless it tells people both what they want to know and how well they can know it. Deciding what to compare is inherently difficult because any two risks differ in many ways. Risk assessors will naturally gravitate toward presenting statistical measures of harm rather than comparing other dimensions of risk that may have more influence on individual and public judgment (e.g., citizens may rather save fewer lives by spending more on Superfund sites than on radon abatement because they perceive the former as also redressing an injustice committed in the past). This focus on risk estimates need not be counterproductive, as long as analysts and regulators understand that risk statistics are like the proverbial lamp post: if the lost keys are underneath it one need not look further, but one should not be surprised not to find them there.

In considering how to compare risk statistics, on the other hand, it is only slightly more difficult to do it well than to do it badly. At a minimum, analysts should estimate and communicate some measure of the lower and upper bounds of each risk ratio, rather than just a measure of central tendency or a qualitative pronouncement about which risk is definitely "worse." In cases where one risk is almost certainly larger than another, this mode of communication should reinforce the distinction and increase confidence and trust (e.g., risk A is at least 10 times larger than risk B, and may be as much as 500 times larger). In other cases such as the Alar/aflatoxin example, where the lower and upper bounds reveal an ambiguous rank order, this fact should not be hidden, but turned from an adversary into an ally by one simple step: admitting that any rank ordering or any decision that flows from it will not be iron-clad, but will be informed by what the numbers say and what they don't (or cannot yet) say. Point estimates of uncertain risk comparisons offer a simplicity that makes decisions easier but makes wrong decisions well-nigh inevitable. Rather than either blinding ourselves to the numbers or letting the numbers usurp all our power to discern and choose, we should start fresh with Schopenhauer's apt advice: "the value of what one knows is doubled if one confesses to not knowing what one does not know."

REFERENCES

1. Cohen BL, Lee IS. A catalog of risks. Health Phys 36:707–722 (1979).
2. Rothschild's numerate arrogance. Nature 276:429 (1978).
3. Roth E, Morgan MG, Fischhoff B, Lave L, and Bostrom A. What do we know about making risk comparisons? Risk Anal 10:375–392 (1990).
4. Office of Technology Assessment, U.S. Congress. Risks to children in school. Environment Program, OTA (in press).
5. Abelson PH. Reflections on the environment. Science 263:591 (1994).
6. Graham JD, Weiner J. Risk versus risk: resolving tradeoffs in health and environmental protection. Cambridge, MA:Harvard University Press (in press).
7. Wilson R, Crouch EAC. Risk assessment and comparisons: an introduction. Science 236:267–270 (1987).
8. Roberts L. Counting on science at EPA. Science 249:616–618 (1990).
9. Finkel AM. Taking aim at environmental risks: questions of feasibility and desirability. The Geneva Papers on Risk and Insurance 17:343–354 (1992).
10. U.S. EPA. Reducing risk: setting priorities and strategies for environmental protection, SAB-EC-90-021. Washington, DC:Environmental Protection Agency, 1990.
11. Finkel AM, Golding D, eds. Worst things first?

The debate over risk-based national environmental priorities. Washington, DC:Resources for the future, 1994.

12. Main J. The big cleanup gets it wrong. Fortune 123:95–101 (1991).

13. Covello VT, Sandman PM, Slovic P. Risk communication, risk statistics, and risk comparisons: a manual for plant managers. Washington, DC:Chemical Manufacturers Association, 1988.

14. Hornstein D. Reclaiming environmental law: a normative critique of comparative risk analysis. Columbia Law Review 92:501–571 (1992).

15. Slovic P, Kraus N, Covello V. Comment: What *should* we know about making risk comparisons? Risk Anal 10:389–392 (1990).

16. Morgan MG, Henrion M. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. New York:Cambridge University Press, 1990.

17. McKone TE, Bogen KT. Predicting the uncertainties in risk assessment. Environ Sci Technol 25:1674–1681 (1991).

18. National Research Council. Science and judgment in risk assessment. Washington, DC: National Academy Press, 1994.

19. Ames BN, Gold LS. Pesticides, risk, and applesauce. Science 244:755–757 (1989).

20. Ames BN. Letter to Hewitt D (CBS Inc., New York, NY), 29 June 1989.

21. Young FE. Weighing food safety risks. FDA Consumer September:8–13 (1989).

22. Hageman FJ. Letter to Finkel AM, 5 August 1991.

23. Gold LS, Stone TH, Stern BR, Manley NB, Ames BN. Rodent carcinogens: setting priorities. Science 258:261–265 (1992).

24. Smith K. Alar three years later: science unmasks a hypothetical health scare. New York:American Council on Science and Health, 1992.

25. Goodman G, Wilson R. Quantitative prediction of human cancer risk from rodent carcinogenic potencies: a closer look at the epidemiological evidence for some chemicals not definitively carcinogenic in humans. Regulatory Toxicology Pharmacology 14:118–146 (1991).

26. Campbell TC, Chen JS, Liu CB, Li JY, Parpia

B. Nonassociation of aflatoxin with primary liver cancer in a cross-sectional ecological survey in the People's Republic of China. Cancer Res 50:6882–6893 (1990).

27. Pao EM, Fleming KH, Guenther PM, Mickle SJ. Foods commonly eaten by individuals: amount per day and per eating occasion. Home Economics Research Report no. 44, Washington, DC:U.S. Deptartment of Agriculture.

28. Cutchins KJ. Letter to Finkel AM, 28 June 1991.

29. Anon. The nuttiest peanut butter. Consumer Reports 588–591 (Sept 1990).

30. Groth N. Letter to Finkel AM, 25 June 1991.

31. Ball JO. Letter to Berteau PE (California Department of Health Services, Berkeley, CA), 3 April 1989.

32. Crouch E, Wilson R. Regulation of carcinogens. Risk Anal 1:47–57 (1981).

33. Cal EPA. Risk-specific intake levels for the proposition 65 carcinogen aflatoxin. Berkeley, CA:California Environmental Protection Agency, 1991.

34. Yeh FS, Yu MC, Mo CC, Luo S, Tong MJ, Henderson BE. Hepatitis B virus, aflatoxins, and hepatocellular carcinoma in Southern Guangxi, China. Cancer Res 49:2506–2509 (1989).

35. Guess H, Crump K, Peto R. Uncertainty estimates for low-dose-rate extrapolations of animal carcinogenicity data. Cancer Research 37:3475–3483 (1977).

36. Portier C, Hoel D. Low-dose-rate extrapolations using the multistage model. Biometrics 39:897–906 (1983).

37. Sielken RL. GEN.T: a general tool for incorporating cancer dose-response extrapolation techniques into quantitative risk assessments, User's manual, version 1. Bryan, TX:Sielken, Inc., 1988.

38. Finkel AM. Computing uncertainty in carcinogenic potency: a bootstrap approach incorporating bayesian prior information. Report to the Office of Policy, Planning, and Evaluation, U.S. EPA. Washington, DC:Resources for the Future, 1988.

39. Goldenthal EI. Two-year oncogenicity study in

[CD-1] mice with UDMH. Unpublished report prepared by International Research and Development Corp., submitted to U.S. EPA by Uniroyal Chemical Co., report #IRDC 399-065, MRID # 413780-01, 1990.

40. Cal EPA. Risk-specific intake level for the proposition 65 carcinogen 1,1-dimethylhydrazine (UDMH). Berkeley, CA:California Environmental Protection Agency, 1992.

41. Toth B. 1,1-Dimethylhydrazine (unsymmetrical) carcinogenesis in mice: light microscopic and ultrastructural studies on neoplastic blood vessels. J Nat Cancer Inst 50:181–187 (1973).

42. Marshall E. A is for apple, Alar, and . . . alarmist? Science 254:20–22 (1991).

43 Kimm VJ. Alar's risks. Science 254:1276 (1991).

44. Finkel AM. Alar: the aftermath. Science 255:664–665 (1992).

45. Finley B, Proctor D, Scott P, Price P, Harrington N, Paustenbach D. Recommended distributions for exposure factors frequently used in health risk assessment. Risk Anal 14:533–552 (1994).

46. Freund JE. Mathematical statistics. NJ:Prentice Hall, 1971.

47. Finkel AM. A quantitative estimate of the extent of human susceptibility to cancer and its implications for risk management. In: Low-dose extrapolation of cancer risks: issues and perspectives (Farland W, Olin S, Park C, Rhomberg L, Scheuplein R, Starr T, Wilson J, eds). Washington, DC:International Life Sciences Institute Press (in press).

48. Peto R, Gray R, Brantom P, Grasso P. Dose and time relationships for tumor induction in the liver and esophagus of 4,080 inbred rats by chronic ingestion of N-nitrosodiethylamine or N-nitrosodimethylamine. Cancer Res 51:6452–6469 (1991).

49. NRDC. Intolerable risk: pesticides in our children's food, Washington, DC:Natural Resources Defense Council,1989.

50. Gold LS. Letter to Finkel AM, 31 July 1994.

51. Ames BN. Gold LS. Carcinogens and human health. Science 251:607–608 (1991).

52. Bell D. Regret in decision-making under uncertainty. Operations Res 30:961–981 (1982).